Supplementary Information

Pavlovian-to-Instrumental Transfer in Intertemporal Choice

Floor Burghoorn, Vivian R. Heuvelmans, Anouk Scheres, Karin Roelofs, and Bernd Figner

Contents

S1. Additional Tasks and Questionnaires

In addition to the tasks described in the main text, participants completed several other tasks or questionnaires that were not analysed for our study. These are listed below.

Post-PIT Liking Ratings — Pairwise Comparisons

In addition to rating each of the Pavlovian cues (coloured squares) individually, participants performed a task in which they were presented with two cues at the same time. In Experiment 1, they were asked to rate how much they liked one cue relative to the other cue, on a scale from 0-100 (with 0 indicating a complete preference for the left cue, and 100 indicating a complete preference for the right cue). In Experiment 2, the task was adjusted so that participants were asked to choose which of the two cues they liked better (left/right) or whether they had no preference. The data from these tasks were not analysed.

Pavlovian Contingency Test — Uncertainty Questions

As described in the main text, participants were tested on their Pavlovian cue-outcome contingency awareness through several multiple-choice questions (e.g., "How many cents belong to this picture?"). Each multiple-choice question was followed by a question asking participants to rate how certain they were of their answer on a 5-point scale, ranging from "very certain" to "not at all certain". Answers to these questions were not analysed.

Other Tasks / Questionnaires

Participants of Experiment 1 completed several additional tasks and questionnaires that were administered for possible future exploratory purposes, but were beyond the scope of the current research. These were a standard intertemporal choice task (Figner et al., 2010); a digit span assessment (Wechsler, 2008); the Risk-Taking and Time Discounting items from the Preference Survey Module (Falk et al., 2016); the Behavioral Activation and Inhibition Scale (Carver & White, 1994); the Barratt Impulsiveness Scale (Patton et al., 1995), and several history of substance use questions. These tasks and questionnaires were not administered in Experiment 2.

S2. Delay Discounting Task Figures

Below, we provide a visual depiction of the delay discounting tasks used to derive the two indifference pairs in Experiment 1 and 2.

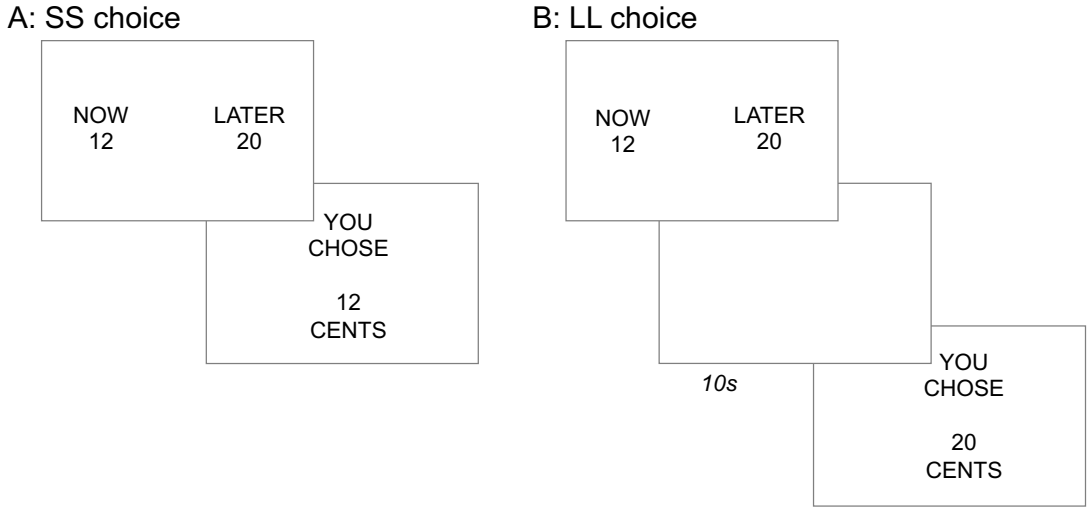

Figure S2.1: Example trial of the delay discounting task in Experiment 1. In each trial of the first task block, participants were presented with a choice between an immediate reward varying between 0 and 20 cents (sooner-smaller reward; SS, e.g., 12 cents), versus a constant reward of 20 cents delayed by 10 seconds (later-larger reward; LL). Participants indicated their choice using the keyboard keys. If they chose the SS, the outcome was presented on the screen immediately (panel A). If they chose the LL, they first experienced a blank screen for 10 seconds (i.e., a real, experienced delay), after which the outcome was presented on the screen (panel B). No explicit information was provided on the exact duration of the delay. All immediate reward amounts, i.e., all integer amounts between 0-20 cents, were presented in combination with the LL twice, in random order, resulting in 42 trials. The indifference value was subsequently derived by fitting a generalized linear model with choice (SS vs LL) as dependent variable and the SS reward amount as predictor. The SS for which the predicted probability of choosing the LL was closest to 0.5 (i.e., closest to indifference) was selected as the immediate medium member of an indifference pair with a delayed large reward of 20 cents in 10 seconds. The second task block served to derive a second indifference pair. It was identical to the first task block, except that the amount of the immediate medium reward derived in the first task block (e.g., 14 cents) was now used as the LL amount. For instance, if this reward amount was 14 cents, each trial in the second block presented a choice between an immediate reward varying between 0 and 14 cents immediately, versus 14 cents in 10 seconds. This task block resulted in an immediate small reward (e.g., 6 cents now) that formed an indifference pair with the delayed medium (e.g., 14 cents in 10 seconds) reward. After the PIT task, we again administered a delay discounting task; the choices presented there were identical to those presented in the pre-PIT delay discounting task, except that

each choice was presented only once, and all choices were presented in one block in an intermixed manner.

Figure S2.2: Example trial of the delay discounting task in Experiment 2. Instead of presenting the choices one by one (as in Experiment 1), we used a titrator that presented all choice options in rows. In the first task block, participants were presented with a series of choices between an immediate reward (SS) ranging from €2 to €20, increasing by steps of €2, versus a constant delayed reward of €20 in 50 days (LL). The indifference value was computed as the mean of the smallest SS that the participant preferred over the LL, and the SS that preceded this SS (i.e., the SS on the preceding row). Thus, in the example above, the indifference value would be €11. This resulted in an indifference pair formed by an immediate medium reward (e.g., €11 now) versus a delayed large reward (€20 in 50 days). The second task block served to derive a second indifference pair. This block was identical to the first task block, except that the amount of the immediate medium reward derived in the first task block was now used as LL amount. For instance, if this reward amount was €11, the titrator in the second block presented a series of choices between an SS ranging from €1-11, increasing by steps of €2, versus a constant LL of €11 in 50 days. This task block was used to derive an immediate small reward (e.g., €6 now) that formed an indifference pair with the delayed medium (e.g., €11 in 50 days) reward. The section titled Deriving Pavlovian Rewards in the main text describes how we dealt with participants whose choices did not result in one straightforward switch point, in accordance with our preregistration. After the PIT task, we again administered a delay discounting task; this task was identical to the pre-PIT delay discounting task, except that the rewards had been increased by €1 to reduce memory effects and prevent participants from simply repeating their choices from the pre-PIT delay discounting task.

S3. Statistical Model Specifications

This section specifies all statistical models of which the results are reported in detail in the main text of the paper. The descriptions specify the dependent variable, the fixed and random effects structure, and response distribution used. More technical details, such as the model syntax and the number of chains and iterations used to run the models in brms, can be found in the R code available online (https://osf.io/7zcsy/). Please note that the statistical models of analyses reported in detail in the supplementary material are specified in the respective supplementary section.

Instrumental Conditioning

Instrumental learning was assessed using a model with response (correct/incorrect) as dependent variable (modelled with a Bernoulli distribution), a fixed intercept, trial type (go/no-go trial), trial number (continuous predictor), and their interaction as fixed effects. The random effects included participants and instrumental stimuli (mushrooms) as grouping variables, with a random intercept, random slopes of both predictors and their interaction varying over both grouping variables, and all random correlations.

Pavlovian Conditioning — Post-PIT Liking Ratings

We assessed Pavlovian conditioning effects using a model with post-PIT liking ratings as dependent variable (modelled with a Gaussian distribution), a fixed intercept, amount (small/medium/large), delay (immediate/delayed), and their interaction as fixed effects. Random effects included participants and Pavlovian stimuli as grouping variables, with a random intercept and random slopes of amount and delay (omitting their interaction) varying over participants, a random intercept and random slopes of amount, delay, and their interaction varying over Pavlovian stimuli, and all random correlations. The random slope of the interaction between amount and delay varying over participants was omitted because our data contained only one observation per cell; including it as random slope would thus make the model unidentifiable.

Transfer Phase — Transfer Test

To test our main transfer hypothesis, we ran a model with response (go/no-go) as dependent variable (modelled with a Bernoulli distribution), a fixed intercept, and amount (small/medium/large), delay (immediate/delayed), trial type (go/no-go) and their interactions as fixed effects. Random effects included participants, instrumental and Pavlovian stimuli as grouping variables, with a random intercept, random slopes of amount, delay, trial type, and their interactions varying over these grouping variables, and all random correlations.

Post-PIT Delay Discounting Task

To investigate the stability of delay discounting from the pre- to the post-PIT delay discounting task, we ran a model with discount rate as dependent variable (modelled with a lognormal distribution), a fixed intercept, and time (pre-/post-PIT), pair (1/2; with 1 indicating the immediate large versus delayed medium reward pair), and their interaction as fixed effects. Note that a constant value of 1 was added to the discount rates to prevent zero-values, allowing us to model the discount rates with a lognormal distribution. Random effects included a random intercept, random slopes of time and pair (omitting their interaction) varying over participants, and all random correlations. The random slope of the interaction between time and pair was omitted because our data contained only one observation per cell, due to which including it as random slope would make the model unidentifiable.

S4. Robustness Check: Pavlovian Reward Values

For some participants, the Pavlovian reward values were not derived directly from their choices in the pre-PIT delay discounting task. To assess the robustness of our results, we reran our main Pavlovian and transfer models with a subsample excluding these participants. Below, we only report the differences between the subsample and full sample.

Experiment 1

Seven participants showed a pattern of responding in the pre-PIT delay discounting task we termed *non-discounting.* As described in the main text, we assigned these participants Pavlovian reward values of 20 (large), 16 (medium) and 12 (small) cents. Furthermore, seven participants were assigned medium and small Pavlovian reward values that deviated slightly from their choice-derived indifference values due to a technical error. These deviations varied between 3 cents below (-3) and 4 cents above (+4) their choice-derived indifference values (n = 4: +1 for the medium and small reward; n = 1: +2 for the medium and +1 for the small reward; n = 1: -2 for the medium and -3 for small reward; n = 1: +2 for the medium and +4 for the small reward).

Rerunning our Pavlovian conditioning model on a subsample excluding these 14 participants (remaining n = 36) resulted in findings similar to those observed in the full sample, except that the difference in post-PIT liking ratings between medium and small rewards was no longer significant (b_{MvsS} = 8.47, 95% CI [-1.26, 18.70]). In addition, after rerunning our transfer model, the difference in go-responding between large and medium cues was no longer significant after correcting for multiple comparisons (b_{LvsM} = 0.43, corrected 95% CI [-0.11, 0.96], uncorrected 95% CI [0.02, 0.87]).

One straightforward explanation for the differences in the effect of reward amount between the full sample and subsample results is that they were caused by a loss of power due to the exclusion of 14 out of 50 participants. A loss of power would explain why, for the transfer effect of amount, the credible interval for the difference between large and medium cues was wider for the subsample (see above) compared to the full sample (b_{LvsM} = 0.45; 95% CI [0.11, 0.78]), whereas the point estimate remained similar. The point estimate and credible interval for the Pavlovian amount effect shows less support for this account, however, as the point estimate became slightly smaller in the subsample (see above) compared to the full sample (b_{MvsS} = 10.00, 95% CI [0.83, 19.70]).

Alternatively, the differences between the full sample and subsample results could stem from the discrepancy between choice-derived indifference values and assigned

Pavlovian reward values for the 14 excluded participants. However, we would expect such differences to appear most strongly for the results regarding indifference pairs, since these results rest strongly on the assumption of individually preference-matched reward pairs. Instead, results involving the indifference pairs were consistent between the full sample and subsample. We would expect the difference between large, medium, and small rewards (i.e., the effect of amount) to exist regardless of whether or not these reward values were derived on a per-participant basis. Nevertheless, it may be that the effect of amount was stronger for participants whose assigned small, medium, and large reward values were more dissimilar in amount compared to their choice-derived reward values. This may, for instance, be the case for the participants classified as non-discounters, who had choice-derived reward values of 20, 19 and 18 cents, but were assigned Pavlovian reward values of 20, 16, and 12 cents. These participants may have experienced the difference between large, medium, and small as larger than participants whose Pavlovian values were directly derived from the delay discounting task, and excluding them may have reduced the effect of amount. In Experiment 2, we addressed this issue by using Pavlovian reward values that were directly derived from the delay discounting task for all participants.

Experiment 2

In Experiment 2, nine participants were classified as non-discounters. In contrast to Experiment 1, these participants were not assigned predefined indifference values, as we aimed to use Pavlovian reward values that approximated the participants' choice-derived indifference values more closely. Thus, we directly used the indifference values derived from the delay discounting task as Pavlovian rewards. None of these participants selected the later-larger reward (LL) in trials where the sooner-smaller (SS) and LL reward amount were equal (e.g., €20 now versus €20 in 50 days), and therefore, none of the Pavlovian reward values were equal in amount. However, as a consequence of this strategy, participants classified as non-discounters had Pavlovian reward values that were close in magnitude (18, 19 and 20 euros). Moreover, two participants exclusively selected the SS on all trials. Due to their choice behaviour, we could not derive the small Pavlovian reward value from their choices, and they were therefore assigned a small reward value of €2. As a robustness check, we reran our main Pavlovian and transfer models while excluding the nine non-discounters and the two participants for whom we could not derive a small reward (remaining $n = 60$). We found no differences regarding significant/non-significant effects compared to the full sample model reported in the main text.

S5. Moderation by Instrumental Accuracy

On average, participants showed satisfactory accuracy at the end of the instrumental conditioning phase in both experiments. However, substantial inter-individual differences in accuracy were observed, raising the question whether this may have influenced the transfer results. Perhaps participants who were uncertain about which instrumental action to perform were more susceptible to the influence of task-irrelevant Pavlovian cues compared to participants who were certain about the appropriate instrumental action. Supporting this notion, non-human animal research has found stronger PIT effects after a period of instrumental extinction (Cartoni et al., 2016; Holmes et al., 2010), and a study in humans found stronger PIT effects in a task version with low (versus high) instrumental reward probability, which increases the uncertainty about the instrumental response (Cartoni et al., 2015). These findings have been explained by proposing that in these situations, the Pavlovian cues increase reward expectancy when the current expectancy is low (due to, e.g., extinction, low reward probability, or poor performance). In other words, the cues provide decision-makers with a straw to base their choice on, which may be most welcome in situations of high uncertainty.

To explore this possibility, we investigated whether the transfer effects of amount, delay, and the indifference pairs were moderated by instrumental accuracy. We used the proportion of correct responses during the final 10 trials of the instrumental conditioning phase as measure of instrumental accuracy, standardized this variable, and added it as linear predictor to our main transfer model. In addition to the variables of our main transfer model (see S2), this model included the instrumental accuracy predictor as a fixed effect (interacting with all other variables) and as a random effect varying over instrumental stimuli (interacting with all other variables) and over Pavlovian stimuli (interacting with all other variables).

Experiment 1

Running the model specified above showed no statistically significant interactions between instrumental accuracy and the transfer effect of amount ($b_{accuracy*LvsS} = 0.12$, 95% CI [-0.49, 0.68]; $b_{accuracy*LvsM} = -0.01$, 95% CI [-0.49, 0.68]; $b_{accuracy*MvsS} = 0.12$, 95% CI [-0.41, 0.67]), the transfer effect of delay ($b_{accuracy*DvsI} = -0.003$, 95% CI [-0.58, 0.61]), or the transfer effect of indifference pair cues ($b_{accuracy*DLvsIM} = -0.001$, 95% CI [-0.79, 0.78]; $b_{accuracy*DMvsIS} = 0.09$, 95% CI [-0.74, 0.95]).

Experiment 2

Although the interaction estimates were somewhat larger compared to Experiment 1, we again observed no statistically significant interactions between instrumental accuracy and the transfer effect of amount ($b_{accuracy*LvsS}$ = -0.35, 95% CI [-0.81, 0.15]; $b_{accuracy*LvsM}$ = -0.16, 95% CI [-0.51, 0.21]; $b_{accuracy*MvsS}$ = -0.19, 95% CI [-0.75, 0.38]), the transfer effect of delay ($b_{accuracy*DvsI}$ = -0.02, 95% CI [-0.33, 0.33]), or the transfer effect of indifference pair cues ($b_{accuracy*DLvsIM}$ = -0.32, 95% CI [-0.90, 0.27]; $b_{accuracy*DMvsIS}$ = -0.12, 95% CI [-0.81, 0.54]).

Conclusions

In summary, we found no evidence for the moderation of the hypothesized transfer effects by instrumental accuracy. These results were consistent across both experiments.

S6. Figures Transfer Effects Per Trial Type

The figures below display the transfer effects as observed in Experiment 1, Experiment 2, and the pooled data, separated for go and no-go trials. Each avers displays the probability of giving a go response, *p*(go) as a function of the amount and delay associated with the Pavlovian cues presented in the background, separated by trial type (go / no-go trials). Model-based means and 95% CIs are displayed.

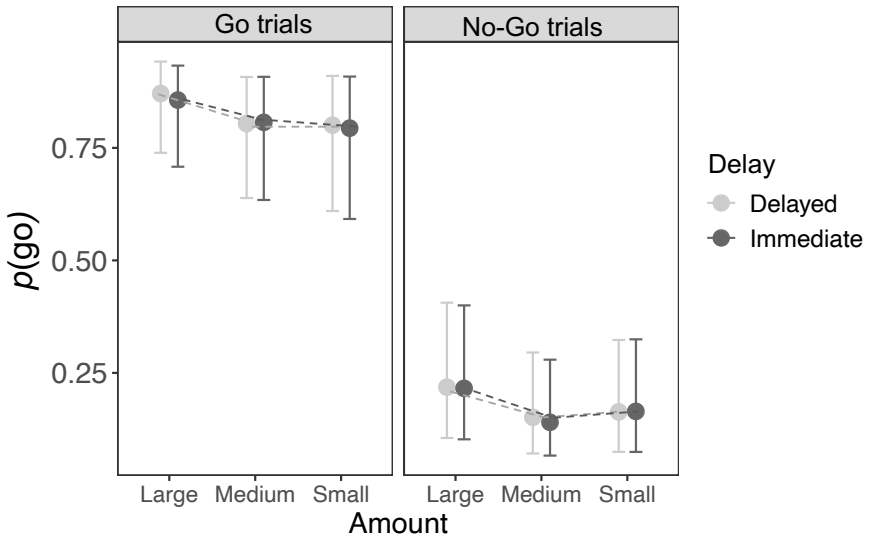

Figure S6.1: Transfer effects separated by trial type in Experiment 1.

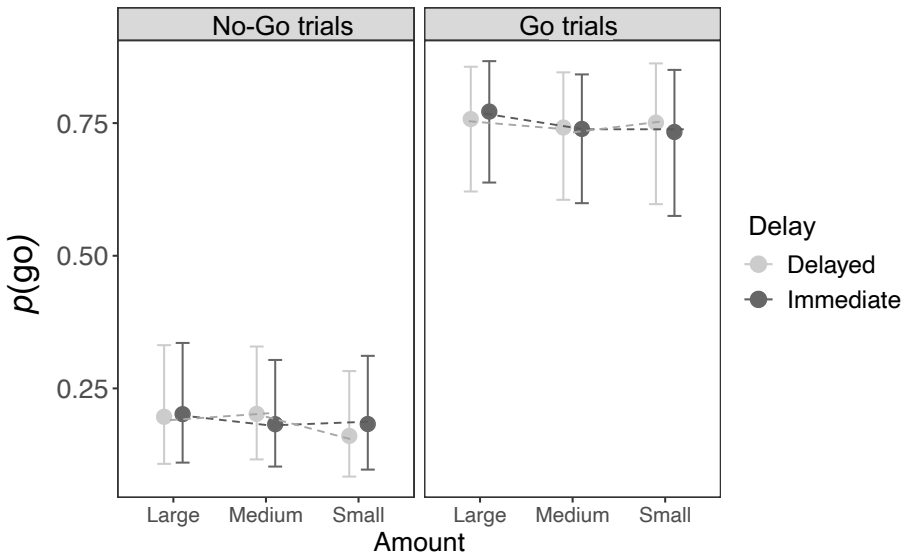

Figure S6.2: Transfer effects separated by trial type in Experiment 2.

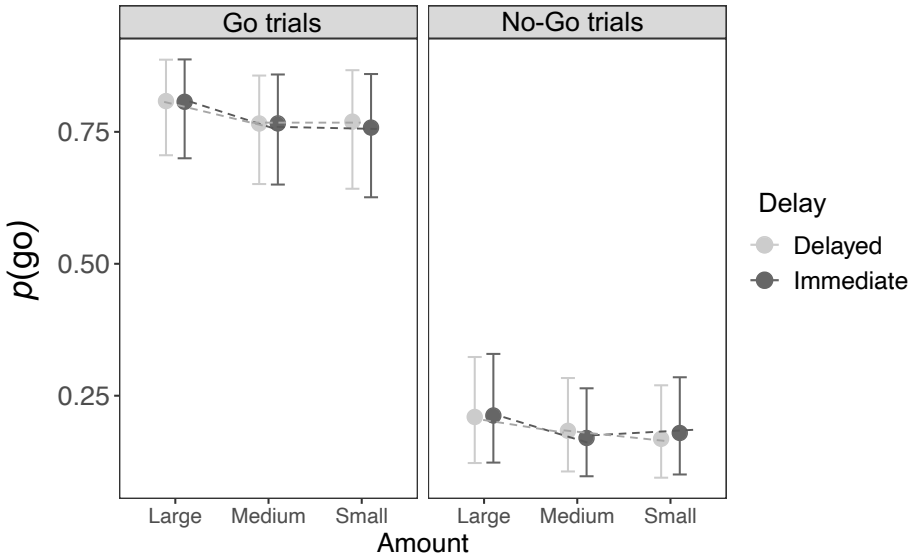

Figure S6.3: Transfer effects separated by trial type in the pooled data.

S7. Planned Comparisons Pavlovian and Transfer Effects

The tables below contain the estimates and 95% CIs of all preregistered planned comparisons for the Pavlovian conditioning and transfer effects in Experiment 1, Experiment 2, and the pooled data. Significant differences are displayed in boldface. Note that estimates in tables involving transfer effects (S7.4 – S7.6) are reported on the log odds scale.

Table S7.1

Planned Comparisons Pavlovian Conditioning Effects — Experiment 1

Contrast			*b*	95% CI
Amount (aggregated across Delay)				
L	-	M	**15.10**	[6.58, 22.40]
	-	S	**25.10**	[13.97, 36.50]
M	-	S	**10.00**	[0.83, 19.70]
Amount (per Delay)				
L del	-	M del	**14.79**	[3.87, 26.20]
	-	S del	**26.67**	[11.69, 42.60]
M del	-	S del	11.89	[-1.45, 26.80]
L imm	-	M imm	**15.36**	[4.28, 26.60]
	-	S imm	**23.55**	[9.58, 39.50]
M imm	-	S imm	8.19	[-4.94, 22.40]
Delay (aggregated across Amount)				
del	-	imm	-6.70	[-14.10, 0.96]
Delay (per Amount)				
L del	-	L imm	-5.85	[-18.00, 7.67]
M del	-	M imm	-5.28	[-18.20, 6.96]
S del	-	S imm	-8.98	[-22.40, 5.10]
Indifference pairs				
L del	-	M imm	9.51	[-3.17, 22.61]
M del	-	S imm	2.91	[-9.40, 16.33]

Table S7.2

Planned Comparisons Pavlovian Conditioning Effects — Experiment 2

Contrast			*b*	95% CI
Amount (aggregated across Delay)				
L	-	M	5.04	[-1.20, 11.00]
	-	S	**18.81**	[11.06, 27.30]
M	-	S	**13.78**	[6.42, 22.00]
Amount (per Delay)				
L del	-	M del	3.34	[-5.12, 11.60]
	-	S del	**17.67**	[6.63, 28.50]
M del	-	S del	**14.33**	[2.38, 25.10]
L imm	-	M imm	6.73	[-1.99, 15.00]
	-	S imm	**19.95**	[8.50, 30.40]
M imm	-	S imm	**13.22**	[2.76, 24.20]
Delay (aggregated across Amount)				
del	-	imm	**-7.63**	[-13.00, -1.59]
Delay (per Amount)				
L del	-	L imm	**-9.52**	[-18.70, -0.44]
M del	-	M imm	-6.13	[-15.90, 3.38]
S del	-	S imm	-7.24	[-18.10, 4.21]
Indifference pairs				
L del	-	M imm	-2.79	[-12.75, 6.06]
M del	-	S imm	7.09	[-2.32, 17.57]

Table S7.3

Planned Comparisons Pavlovian Conditioning Effects — Pooled Data

Contrast			*b*	95% CI
Amount (aggregated across Delay)				
L	-	M	**9.99**	[5.76, 14.40]
	-	S	**21.95**	[15.91, 28.30]
M	-	S	**11.96**	[6.91, 17.20]
Amount (per Delay)				
L del	-	M del	**8.79**	[2.75, 15.00]
	-	S del	**21.98**	[14.18, 30.00]
M del	-	S del	**13.19**	[5.53, 20.50]
L imm	-	M imm	**11.19**	[4.95, 16.90]
	-	S imm	**21.93**	[14.08, 30.20]
M imm	-	S imm	**10.73**	[3.11, 18.10]
Delay (aggregated across Amount)				
del	-	imm	**-7.25**	[-12.90, -1.95]
Delay (per Amount)				
L del	-	L imm	**-8.03**	[-15.70, -0.72]
M del	-	M imm	-5.63	[-13.10, 2.13]
S del	-	S imm	-8.09	[-16.30, 0.53]
Indifference pairs				
L del	-	M imm	3.16	[-4.20, 10.89]
M del	-	S imm	5.10	[-2.92, 12.82]

Table S7.4

Planned Comparisons Transfer Effects — Experiment 1

Contrast			Aggregated across Trial Types		Go Trials		No-Go Trials	
			b	95% CI	*b*	95% CI	*b*	95% CI
Amount (aggregated across Delays)								
L	-	M	**0.45**	[0.11, 0.78]	0.42	[-0.01, 0.85]	**0.47**	[0.04, 0.89]
	-	S	0.44	[-0.07, 0.96]	0.50	[-0.14, 1.15]	0.37	[-0.27, 0.99]
M	-	S	-0.01	[-0.42, 0.43]	0.08	[-0.48, 0.64]	-0.09	[-0.48, 0.64]
Amount (per Delay)								
L del	-	M del	**0.47**	[0.01, 0.95]	0.50	[-0.10, 1.14]	0.44	[-0.16. 1.05]
	-	S del	0.47	[-0.23, 1.15]	0.57	[-0.30, 1.45]	0.38	[-0.45, 1.25]
M del	-	S del	-0.003	[-0.59, 0.62]	0.07	[-0.75, 0.84]	-0.08	[-0.87, 0.69]
L imm	-	M imm	0.42	[-0.06, 0.88]	0.35	[-0.26, 0.95]	0.49	[-0.11, 1.08]
	-	S imm	0.40	[-0.31, 1.08]	0.43	[-0.43, 1.32]	0.38	[-0.46, 1.24]
M imm	-	S imm	-0.01	[-0.66, 0.57]	0.09	[-0.72, 0.90]	-0.11	[-0.88, 0.70]
Delay (aggregated across Amounts)								
del	-	imm	0.05	[-0.41, 0.49]	0.07	[-0.46, 0.64]	0.03	[-0.50, 0.59]
Delay (per Amount)								
L del	-	L imm	0.09	[-0.54, 0.73]	0.17	[-0.66, 0.94]	0.02	[-0.71, 0.84]
M del	-	M imm	0.03	[-0.57, 0.65]	0.01	[-0.74, 0.81]	0.06	[-0.69, 0.84]
S del	-	S imm	0.02	[-0.72, 0.72]	0.03	[-0.83, 0.99]	0.02	[-0.86, 0.90]
Indifference pairs								
L del	-	M imm	0.51	[-0.11, 1.15]	0.51	[-0.29, 1.27]	0.51	[-0.28, 1.29]
M del	-	S imm	0.02	[-0.62, 0.68]	0.09	[-0.77, 0.88]	-0.05	[-0.83, 0.80]

Table S7.5

Planned Comparisons Transfer Effects — Experiment 2

Contrast			Aggregated across Trial Types		Go Trials		No-Go Trials	
			b	95% CI	*b*	95% CI	*b*	95% CI
Amount (aggregated across Delays)								
L	-	M	0.09	[-0.17, 0.34]	0.13	[-0.20, 0.48]	0.04	[-0.30, 0.39]
	-	S	0.15	[-0.24, 0.53]	0.12	[-0.38, 0.69]	0.18	[-0.34, 0.69]
M	-	S	0.06	[-0.26, 0.39]	-0.01	[-0.45, 0.41]	0.14	[-0.29, 0.59]
Amount (per Delay)								
L del	-	M del	0.03	[-0.34, 0.40]	0.08	[-0.42, 0.56]	-0.03	[-0.52, 0.47]
	-	S del	0.14	[-0.42, 0.70]	0.03	[-0.69, 0.73]	0.25	[-0.48, 0.98]
M del	-	S del	0.11	[-0.37, 0.60]	-0.05	[-0.67, 0.58]	0.28	[-0.36, 0.93]
L imm	-	M imm	0.15	[-0.21, 0.52]	0.18	[-0.31, 0.66]	0.12	[-0.37, 0.60]
	-	S imm	0.16	[-0.39, 0.71]	0.21	[-0.50, 0.94]	0.12	[-0.60, 0.85]
M imm	-	S imm	0.01	[-0.46, 0.50]	0.02	[-0.60, 0.66]	0.0002	[-0.63, 0.66]
Delay (aggregated across Amounts)								
del	-	imm	-0.003	[-0.36, 0.31]	0.01	[-0.38, 0.40]	-0.02	[-0.41, 0.38]
Delay (per Amount)								
L del	-	L imm	-0.05	[-0.57, 0.45]	-0.08	[-0.71, 0.55]	-0.03	[-0.67, 0.61]
M del	-	M imm	0.07	[-0.39, 0.55]	0.02	[-0.57, 0.59]	0.13	[-0.47, 0.70]
S del	-	S imm	-0.03	[-0.60, 0.58]	0.10	[-0.62, 0.83]	-0.15	[-0.87, 0.61]
Indifference pairs								
L del	-	M imm	0.10	[-0.37, 0.57]	0.10	[-0.47, 0.70]	0.09	[-0.49, 0.69]
M del	-	S imm	0.09	[-0.43, 0.58]	0.04	[-0.59, 0.67]	0.13	[-0.54, 0.74]

Table S7.6

Planned Comparisons Transfer Effects — Pooled Data

Contrast			Aggregated across Trial Types		Go Trials		No-Go Trials	
			b	95% CI	*b*	95% CI	*b*	95% CI
Amount (aggregated across Delays)								
L	-	M	**0.27**	[0.06, 0.48]	**0.28**	[0.01, 0.56]	0.26	[-0.01, 0.54]
	-	S	0.28	[-0.05, 0.59]	0.29	[-0.14, 0.70]	0.26	[-0.17, 0.65]
M	-	S	0.01	[-0.27, 0.27]	0.01	[-0.34, 0.37]	-0.003	[-0.34, 0.36]
Amount (per Delay)								
L del	-	M del	0.25	[-0.03, 0.54]	0.30	[-0.08, 0.67]	0.21	[-0.17, 0.59]
	-	S del	0.28	[-0.14, 0.68]	0.27	[-0.29, 0.78]	0.28	[-0.25, 0.83]
M del	-	S del	0.02	[-0.35, 0.40]	-0.02	[-0.54, 0.45]	0.07	[-0.43, 0.57]
L imm	-	M imm	**0.29**	[0.01, 0.58]	0.26	[-0.11, 0.65]	0.31	[-0.06, 0.69]
	-	S imm	0.28	[-0.14, 0.69]	0.32	[-0.22, 0.85]	0.23	[-0.31, 0.78]
M imm	-	S imm	-0.01	[-0.39, 0.40]	0.05	[-0.47, 0.54]	-0.08	[-0.58, 0.44]
Delay (aggregated across Amounts)								
del	-	imm	0.02	[-0.25, 0.28]	0.03	[-0.27, 0.24]	0.01	[-0.29, 0.32]
Delay (per Amount)								
L del	-	L imm	0.01	[-0.37, 0.35]	0.02	[-0.43, 0.46]	-0.01	[-0.47, 0.44]
M del	-	M imm	0.04	[-0.34, 0.40]	-0.01	[-0.46, 0.46]	0.09	[-0.37, 0.58]
S del	-	S imm	0.004	[-0.41, 0.44]	0.07	[-0.47, 0.59]	-0.06	[-0.57, 0.49]
Indifference pairs								
L del	-	M imm	0.29	[-0.07, 0.67]	0.29	[-0.21, 0.73]	0.30	[-0.16, 0.74]
M del	-	S imm	0.03	[-0.36, 0.41]	0.04	[-0.44, 0.54]	0.01	[-0.49, 0.49]

S8. ROPE Tests Indifference Pairs

As described in the main text, and as expected, we found no significant difference in post-PIT liking ratings between cues associated with members of the indifference pairs, nor did we find a significant difference in instrumental responding in the transfer phase between such indifference pair cues. However, not finding a significant difference does not provide statistical support for the null hypothesis, i.e., it does not imply that the liking ratings and the proportion of instrumental go-responding between these cues were equivalent. Therefore, we further explored the equivalence in liking ratings and instrumental responding between these cues.

We had originally preregistered to do this by examining at which credible interval level the CI excludes 0, with a lower CI level indicating more support for equivalence, as this implies that a smaller CI is required for the posterior distribution to exclude 0. Although this approach can be informative, it does not take into account the shape of the posterior distribution, such that for very wide posterior distributions—which are not strongly centred around the null value and therefore do not strongly support the null value—the CI level at which 0 is excluded will also be low. This may lead us to erroneously conclude that the ratings or instrumental responses between the indifference pair cues are equivalent. Since we preregistered this approach, we do provide its results below for completeness. However, we additionally performed non-preregistered Region of Practical Equivalence tests (Kruschke, 2018), which do take into account the posterior distribution and which we therefore consider to be more appropriate and informative.

A ROPE test requires defining a Region of Practical Equivalence (ROPE) around the null value, declaring all values inside this region as practically equivalent to the null value, and evaluating how much of the posterior distribution falls inside this region. This can be evaluated in one of two ways. First, as Figures S8.1-8.6 indicate, none of the 95% HDIs associated with the difference between indifference pair cues fell *entirely* within the ROPE. Thus, according to the decision criterion proposed by (Kruschke, 2018), we cannot conclude that the liking ratings or instrumental responses between indifference pair cues were equivalent. In addition, however, we provide a more quantitative index of equivalence by evaluating the *proportion* of the whole posterior that fell inside the ROPE, with a higher proportion indicating more support for equivalence. In interpreting this index, we labelled proportions of 0-50% falling inside the ROPE as weak support for equivalence, 50-80% as moderate support, and 80-100% as strong support. We acknowledge that these labels are

somewhat arbitrary, and we therefore also report the numerical proportions, allowing anyone to interpret these values themselves.

As explained in the main text, for the ROPE test on the equivalence of post-PIT liking ratings, we used a ROPE radius based on a Cohen's *d* value of 0.15, translating to ROPEs of 0 ± 4-5 points on a scale from 0-100 (reflecting the difference in ratings between the two indifference pair cues). The exact ROPE radii, which are reported below, differed per ROPE test on liking ratings, because the unstandardized ROPE radius depended on the standard deviation of the observed difference scores.

For the ROPE test on the equivalence of instrumental responding between indifference pair cues, we used a ROPE radius based on Cohen's *d* of 0.10, translating to a radius of 0.5 ± 0.045 on the probability scale. As this radius does not depend on the dispersion of the observed scores, it is identical for each ROPE test on go-responding.

Experiment 1

Pavlovian Conditioning

ROPE. The Cohen's *d* value of 0.15 resulted in a ROPE radius of 4.44 (i.e., a ROPE of 0 ± 4.44) for the first indifference pair (delayed large versus immediate medium reward) and 4.18 (i.e., a ROPE of 0 ± 4.18) for the second indifference pair (delayed medium versus immediate small reward). Figure S8.1A shows that for the first indifference pair, 18% of the posterior distribution fell inside the ROPE. Figure S8.1B shows that for the second indifference pair, 47% falls inside the ROPE. These results suggest that, at least for this ROPE radius, neither of the two pairs show strong support for the null value (i.e. equivalence in post-PIT liking ratings), although more support was found for the second pair. Figure S8.1C and S8.1D show how the proportion of the posterior that falls inside the ROPE as a function of the ROPE radius.

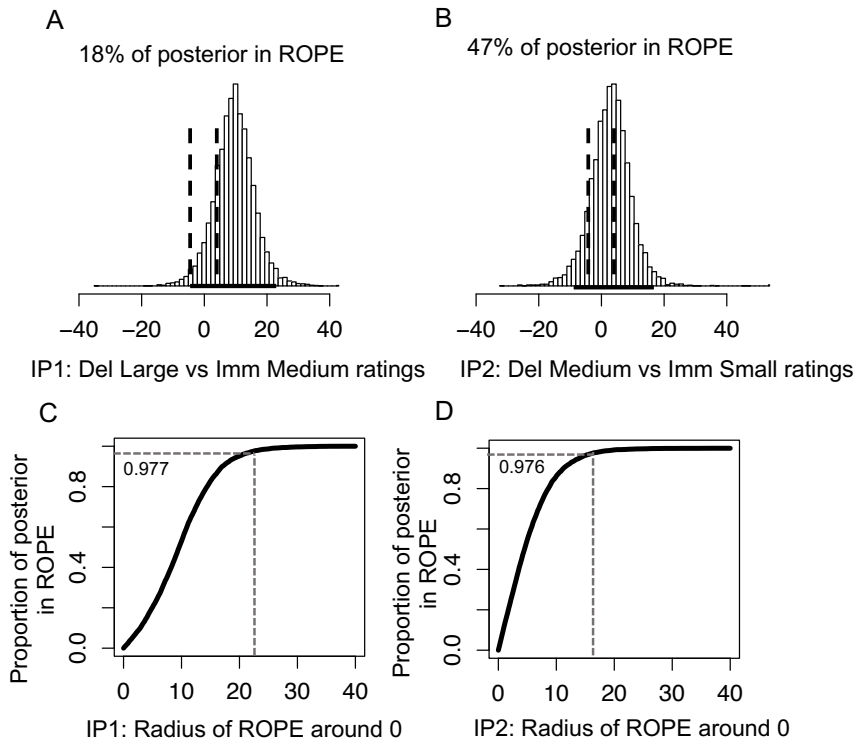

Figure S8.1: Pavlovian ROPE test results of Experiment 1. Panel A-B: Posterior distribution of the difference in post-PIT liking ratings between indifference pairs cues in Experiment 1, for indifference pair 1 (panel A, IP1: delayed large versus immediate medium) and 2 (panel B: IP2: delayed medium versus immediate small). Dashed vertical lines mark the ROPE limits around the null value. The horizontal line in bold marks the 95% Highest Density Interval (HDI). Panel C-D: Proportion of the posterior distribution falling inside the ROPE as a function of the radius (half width) of the ROPE for indifference pair 1 (panel C) and 2 (panel D). ROPE radii represent the difference in post-PIT liking ratings between indifference pair cues. The dashed vertical line marks the ROPE radius at which the 95% HDI falls completely within the ROPE; the dashed horizontal line indicates the proportion of the *whole* posterior distribution that falls within the ROPE for this radius.

Credible Intervals. For the cues associated with the first indifference pair, the CI of the difference in post-PIT liking ratings excluded 0 at an 85% CI level. For the cues associated with the second indifference pair, the CI excluded 0 at a 40% CI level. These results are in line with the ROPE results.

Transfer Effects

ROPE. Figure S8.2A-B show that for the first indifference pair, 11% of the posterior distribution fell inside the ROPE, whereas for the second indifference pair, 45% fell inside the ROPE. Thus, in line with the Pavlovian conditioning results, we observed weak support for

the equivalence in go-responding of both indifference pairs, although we observed more support for the second pair. Figure S8.2C-D show how the proportion of the posterior that falls inside the ROPE as a function of the ROPE radius.

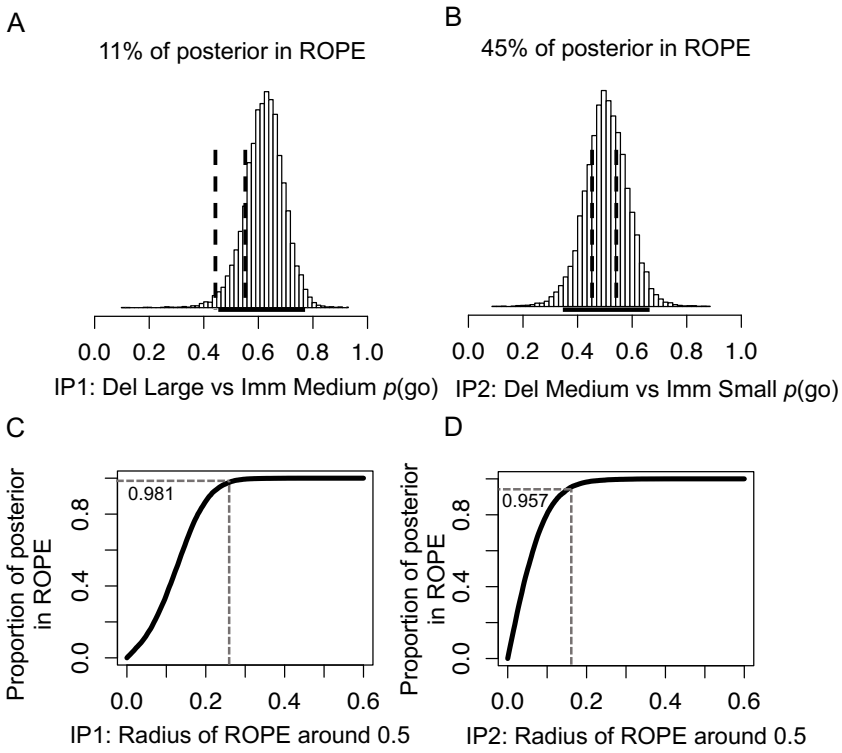

Figure S8.2: Transfer ROPE test results of Experiment 1. Panel A-B: Posterior distribution of go-responding in the presence of Pavlovian cues associated with the delayed versus the immediate member of the indifference pair in Experiment 1, for indifference pair 1 (panel A, IP1: delayed large versus immediate medium) and 2 (panel B: IP2: delayed medium versus immediate small). Values are on the probability scale. Dashed vertical lines mark the ROPE limits around the null value. The bold horizontal lines mark the 95% Highest Density Interval (HDI). Panel C-D: Proportion of the posterior distribution falling inside the ROPE as a function of the radius (half width) of the ROPE, for indifference pair 1 (panel C) and 2 (panel D). ROPE radii are on the probability scale. The dashed vertical line marks the ROPE radius at which the 95% HDI falls completely within the ROPE; the dashed horizontal line indicates the proportion of the *whole* posterior distribution that falls within the ROPE for this radius.

Credible Intervals. The CI of the difference in go-responding excluded 0 at an 85% CI level for the first indifference pair, and at a 5% CI level for the second indifference pair. Similar to the ROPE results, the data thus provide weak support for the equivalence of the cues associated with the first indifference pair. In contrast to the ROPE results, the CIs do suggest

strong support for equivalence of the second pair. However, as described above, we deem the ROPE results to be more informative than the CIs.

Experiment 2

Pavlovian Conditioning

ROPE. Figure 8.3A shows that for a ROPE radius of 4.01 for the first indifference pair, 55% of the posterior distribution fell inside the ROPE, providing moderate support for equivalence. Figure 8.3B shows that for a ROPE radius of 4.02 for the second indifference pair, 24% of the posterior distribution fell inside the ROPE, providing weak support for equivalence. In contrast to Experiment 1, Experiment 2 thus provides most support for the subjective equivalence of the cues associated with the first indifference pair. Figure S8.3C and S8.3D show how the proportion of the posterior that falls inside the ROPE as a function of the ROPE radius.

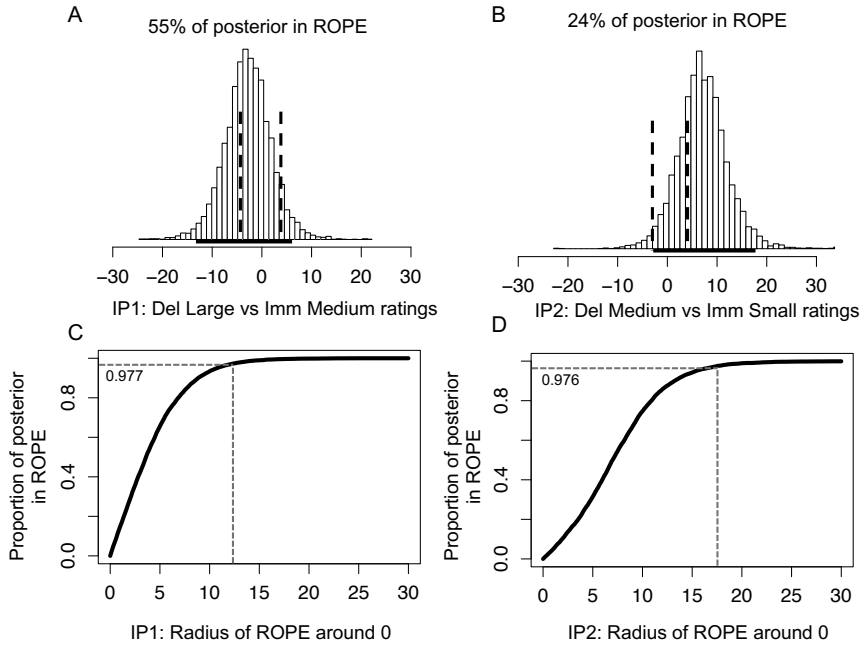

Figure S8.3: Pavlovian ROPE test results of Experiment 2. Panel A-B: Posterior distribution of the difference in post-PIT liking ratings between indifference pairs cues in Experiment 2, for indifference pair 1 (panel A, IP1: delayed large versus immediate medium) and 2 (panel B: IP2: delayed medium versus immediate small). Dashed vertical lines mark the ROPE limits around the null value. The horizontal line in bold marks the 95% Highest Density Interval (HDI). Panel C-D: Proportion of the posterior distribution falling inside the ROPE as a function of the radius (half width) of the ROPE for indifference pair 1 (panel C) and 2 (panel D). ROPE radii represent the difference in post-PIT liking

ratings between indifference pair cues. The dashed vertical line marks the ROPE radius at which the 95% HDI falls completely within the ROPE; the dashed horizontal line indicates the proportion of the *whole* posterior distribution that falls within the ROPE for this radius.

Credible Intervals. For the cues associated with the first indifference pair, the CI of the difference in post-PIT liking ratings excluded 0 at an 45% CI level. For the cues associated with the second indifference pair, the CI excluded 0 at an 80% CI level. These results are in line with the ROPE results.

Transfer Effects

ROPE. Again, the ROPE radius was set at 0.045. Figure S8.4A shows that for the first indifference pair, 55% of the posterior distribution fell inside the ROPE. Figure S8.4B shows that for the second indifference pair, 54% fell inside the ROPE. This indicates a moderate degree of support for the equivalence in go-responding for both indifference pairs, contrasting with the Pavlovian conditioning results. Figure S8.4C and S8.4D show how the proportion of the posterior that falls inside the ROPE as a function of the ROPE radius.

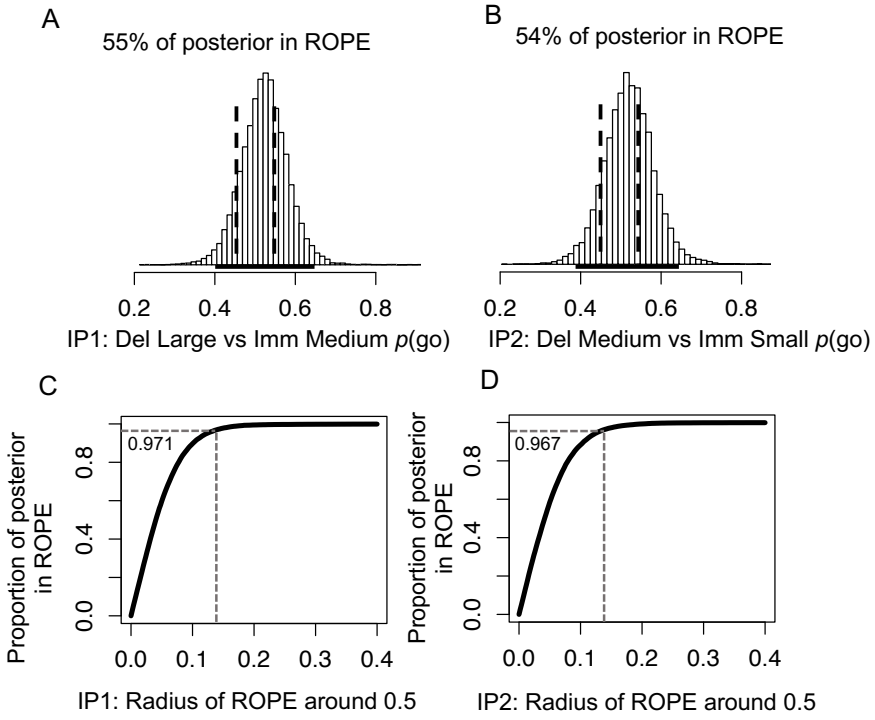

Figure S8.4: Transfer ROPE test results of Experiment 2. Panel A-B: Posterior distribution of go-responding in the presence of Pavlovian cues associated with the delayed versus the immediate

member of the indifference pair in Experiment 2, for indifference pair 1 (panel A, IP1: delayed large versus immediate medium) and 2 (panel B: IP2: delayed medium versus immediate small). Values are on the probability scale. Dashed vertical lines mark the ROPE limits around the null value. The bold horizontal lines mark the 95% Highest Density Interval (HDI). Panel C-D: Proportion of the posterior distribution falling inside the ROPE as a function of the radius (half width) of the ROPE, for indifference pair 1 (panel C) and 2 (panel D). ROPE radii are on the probability scale. The dashed vertical line marks the ROPE radius at which the 95% HDI falls completely within the ROPE; the dashed horizontal line indicates the proportion of the *whole* posterior distribution that falls within the ROPE for this radius.

Credible Intervals. The CI of the difference in go-responding excluded 0 at a 30% CI level for the first indifference pair, and at a 20% CI level for the second indifference pair. Similar to the ROPE results, the CIs thus seem to provide support for the subjective equivalence of both indifference pairs.

Pooled Data Analyses

Pavlovian Conditioning

ROPE. Figure 8.5A shows that for a ROPE radius of 4.28 for the first indifference pair, 60% of the posterior distribution fell inside the ROPE, showing moderate support for the subjective equivalence of the cues. Figure 8.5B shows that for a ROPE radius of 4.08 for the second indifference pair, 37% of the posterior distribution fell inside the ROPE, showing weak support for equivalence. Figure S8.5C and S8.5D show how the proportion of the posterior that falls inside the ROPE as a function of the ROPE radius.

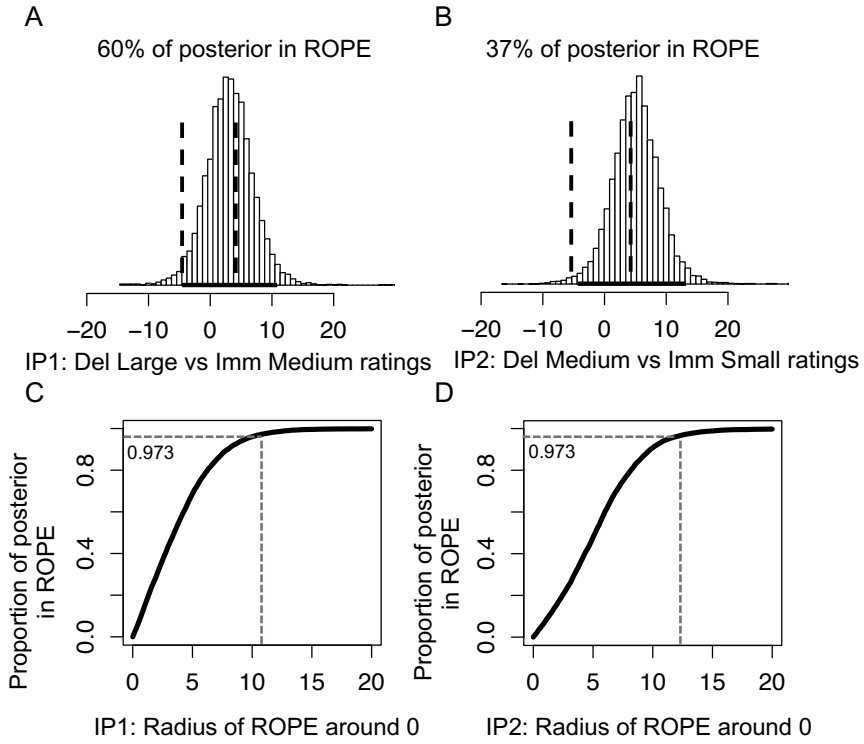

Figure S8.5: Pavlovian ROPE test results of the pooled data. Panel A-B: Posterior distribution of the difference in post-PIT liking ratings between indifference pairs cues in the pooled data, for indifference pair 1 (panel A, IP1: delayed large versus immediate medium) and 2 (panel B: IP2: delayed medium versus immediate small). Dashed vertical lines mark the ROPE limits around the null value. The horizontal line in bold marks the 95% Highest Density Interval (HDI). Panel C-D: Proportion of the posterior distribution falling inside the ROPE as a function of the radius (half width) of the ROPE for indifference pair 1 (panel C) and 2 (panel D). ROPE radii represent the difference in post-PIT liking ratings between indifference pair cues. The dashed vertical line marks the ROPE radius at which the 95% HDI falls completely within the ROPE; the dashed horizontal line indicates the proportion of the *whole* posterior distribution that falls within the ROPE for this radius.

Credible Intervals. For the cues associated with the first indifference pair, the CI of the difference in post-PIT liking ratings excluded 0 at a 60% CI level. For the cues associated with the second indifference pair, the CI excluded 0 at an 80% CI level. In line with the ROPE results, this suggests more support for the subjective equivalence of the cues associated with the first indifference pair.

Transfer Effects

ROPE. The ROPE radius was again set at 0.045. Figure S8.6A shows considerably more support for the equivalence in go-responding between cues associated with the second

compared to the first indifference pair. That is, for the first indifference pair, 25% of the posterior distribution fell inside the ROPE, providing weak support for the null value. Figure S8.6B shows that for the second indifference pair, 68% fell inside the ROPE, providing moderate support for the null value. Figure S8.6C and S8.6D show how the proportion of the posterior that falls inside the ROPE as a function of the ROPE radius.

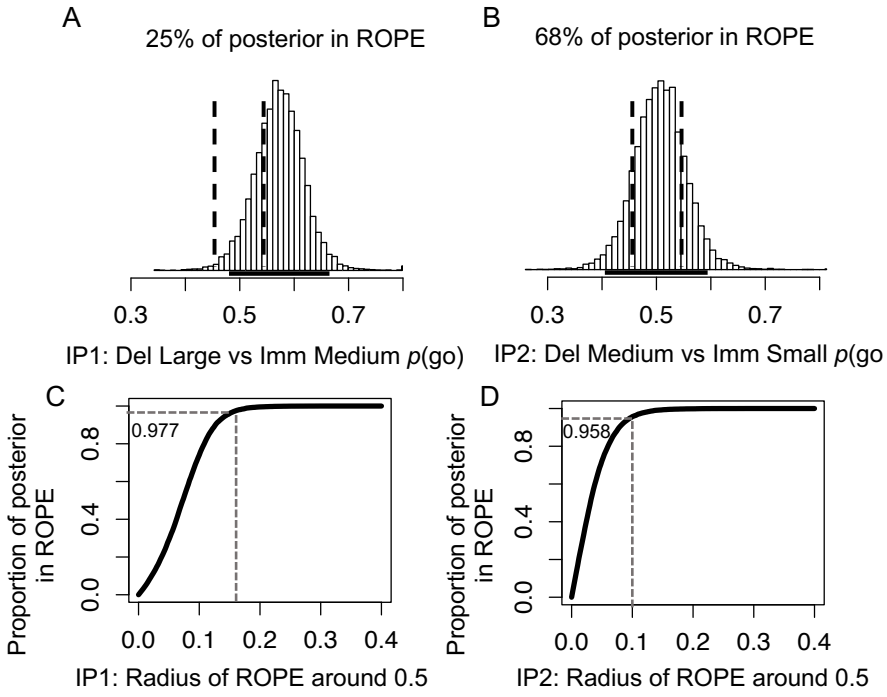

Figure S8.6: Transfer ROPE test results of the pooled data. A-B: Posterior distribution of go-responding in the presence of Pavlovian cues associated with the delayed versus the immediate member of the indifference pair in the pooled data, for indifference pair 1 (panel A, IP1: delayed large versus immediate medium) and 2 (panel B: IP2: delayed medium versus immediate small). Values are on the probability scale. Dashed vertical lines mark the ROPE limits around the null value. The bold horizontal lines mark the 95% Highest Density Interval (HDI). Panel C-D: Proportion of the posterior distribution falling inside the ROPE as a function of the radius (half width) of the ROPE, for indifference pair 1 (panel C) and 2 (panel D). ROPE radii are on the probability scale. The dashed vertical line marks the ROPE radius at which the 95% HDI falls completely within the ROPE; the dashed horizontal line indicates the proportion of the *whole* posterior distribution that falls within the ROPE for this radius

Credible Intervals. The CI of the difference in go-responding excluded 0 at an 85% CI level for the first indifference pair, and at a 10% CI level for the second indifference pair. The

stronger support for equivalence of the second, compared to the first pair, is in line with the ROPE results.

Conclusion

The results of the ROPE tests described above show that the evidence for equivalence in post-PIT liking ratings or go-responding was not consistent across experiments and indifference pairs. Whereas Experiment 1 provided more support for equivalence in indifference pair 2 (delayed large versus immediate medium) than indifference pair 1 (delayed medium versus immediate small), Experiment 2 showed the opposite pattern. When pooled across both experiments, the Pavlovian conditioning data showed more support for equivalence in post-PIT liking ratings for indifference pair 1, and the transfer data showed more support for equivalence in go-responding for indifference pair 2. None of the ROPE tests, however, showed strong support for the equivalence. Thus, although we did not find significant differences in ratings or go-responding between indifference pair cues (as hypothesized), we only found weak to moderate evidence for an actual equivalence in ratings or go-responding.

S9. Pre-PIT Liking Ratings

To investigate whether there were any differences between ratings of the coloured squares (Pavlovian cues) *before* Pavlovian conditioning, which would most likely indicate participants' idiosyncratic colour preferences, we ran a mixed-effects model with pre-PIT liking rating as dependent variable, a fixed intercept, square colour (red / purple / light blue dark blue / orange / green) as fixed effect, and a random intercept varying over participants. Pairwise comparisons were performed to compare all squares' ratings.

Experiment 1

The average rating across all squares was 49.91 on a scale from 0-100 (95% CI [45.97, 53.83]). The light blue square received the highest ratings (M_{LB} = 61.50), and was rated significantly higher than all other squares (b_{LBvsR} = 17.10, 95% CI [10.24, 23.37]; b_{LBvsP} = 12.33, 95% CI [5.44, 18.92]); b_{LBvsDB} = 10.20, 95% CI [3.51, 16.90]); b_{LBvsO} = 15.28, 95% CI [8.57, 21.88]); b_{LBvsG} = 14.20, 95% CI [7.51, 20.96]). The red square received the lowest rating (M_R = 44.40), significantly lower than the light blue square (see above) and the dark blue square (M_{DB} = 51.30; b_{RvsDB} = -6.90, 95% CI [-23.37, -10.24]). No other significant differences between squares were found.

Experiment 2

The average rating across all squares was 56.95 (95% CI [53.22, 60.76]). Again, the light blue square received the highest ratings (M_{LB} = 67.10), and was rated significantly higher than all other squares (b_{LBvsR} = 18.74, 95% CI [13.21, 24.62]; b_{LBvsP} = 10.26, 95% CI [4.67, 16.30]); b_{LBvsDB} = 7.14, 95% CI [1.24, 12.53]); b_{LBvsO} = 9.02, 95% CI [3.57, 15.06]); b_{LBvsG} = 15.47, 95% CI [10.02, 21.40]). Moreover, again, the red square received the lowest rating (M_R = 48.30), and was rated significantly lower than all squares except the green square (b_{RvsP} = -8.49, 95% CI [-14.32, -2.78]; $b_{RvsLB:}$ -18.74, 95% CI [24.62, 13.21]; b_{RvsDB} = -11.61, 95% CI [-17.33, -5.80]; b_{RvsO} = -9.72, 95% CI [-15.30, -3.75]; b_{RvsG} = -3.27, 95% CI [-8.97, 2.44]). Finally, the green stimulus (M_G = 51.60) was rated significantly lower than the light blue (see above), dark blue (M_{DB} = 59.90; b_{GvsDB} = -8.34, 95% CI [-2.54, -13.91]), and orange square (M_O = 58.00, b_{GvsO} = -6.45, 95% CI [-0.91, -12.18]).

Conclusion

The results described above indicate significant differences in pre-PIT liking ratings between the stimuli used as Pavlovian cues, with a consistent preference for the light blue square, and a dislike for the red square. Random assignment of these stimuli to Pavlovian rewards prevented this from confounding the effects of Pavlovian conditioning on post-PIT

liking ratings. In addition, by including square colour as grouping variable in our analyses, we accounted for any potential differences in Pavlovian conditioning or transfer effects caused by pre-existing colour preferences.

S10. Performance Checks Transfer Phase

Prior to testing for the hypothesized transfer effects, we performed two performance checks on the transfer phase data.

Instrumental Accuracy Check

As the transfer phase did not immediately follow the instrumental conditioning phase (with the Pavlovian phase in between), we checked whether participants still showed sufficient instrumental accuracy at the start of the transfer phase. We therefore compared instrumental accuracy during the final 10 trials of the instrumental conditioning phase with the first 10 trials of the transfer phase. Figure 3 and Figure 4 in the main text suggest that average accuracy at the start of the transfer phase was similar to that at the end of the instrumental phase. To statistically test this visual impression, we ran a mixed effects model with response (correct/incorrect) as dependent variable, a fixed intercept, phase (last 10 trials of instrumental phase, iLast / first 10 trials of transfer phase, tFirst), trial type (go/no-go trial) and their interaction as fixed effects. The random effects structure included a random intercept and random slopes of phase, trial type, and their interaction varying over the grouping variables participants and instrumental stimuli (mushrooms), plus all random correlations.

Experiment 1. The model showed no significant effect of phase ($b_{\text{iLastvstFirst}} = 0.08$, 95% CI [-0.48, 0.65]), indicating no significant difference in instrumental accuracy between the end of the instrumental phase ($M_{\text{iLast}} = 0.79$) and start of the transfer phase ($M_{\text{tFirst}} = 0.77$). Furthermore, we found no significant interaction between phase and trial type ($b_{\text{iLastvstFirst*GovsNo-Go}} = 0.27$, 95% CI [-0.96, 1.59]), indicating that the non-significant effect of phase did also not significantly differ between go- and no-go trials. Accordingly, the effect of phase was non-significant in both go ($M_{\text{iLast}} = 0.81$; $M_{\text{tFirst}} = 0.77$; $b_{\text{iLastvstFirst}} = 0.22$, 95% CI [-0.65, 1.05]) and no-go trials ($M_{\text{iLast}} = 0.76$; $M_{\text{tFirst}} = 0.77$; $b_{\text{iLastvstFirst}} = -0.05$, 95% CI [-0.96, 0.77]).

Experiment 2. Similar to Experiment 1, we observed no significant effect of phase ($M_{\text{iLast}} = 0.76$; $M_{\text{tFirst}} = 0.72$; $b_{\text{iLastvstFirst}} = 0.18$, 95% CI [-0.33, 0.64]). Moreover, this effect again did not interact with trial type ($b_{\text{iLastvstFirst*GovsNo-Go}} = 0.35$, 95% CI [-0.96, 1.58]), and the effect of phase was non-significant in both go ($M_{\text{iLast}} = 0.77$; $M_{\text{tFirst}} = 0.70$; $b_{\text{iLastvstFirst}} = 0.35$, 95% CI [-0.45, 1.17]) and no-go trials ($M_{\text{iLast}} = 0.75$; $M_{\text{tFirst}} = 0.75$; $b_{\text{iLastvsTfirst}} = 0.002$, 95% CI [-0.82, 0.74]).

Pavlovian Extinction Check

During the transfer phase, Pavlovian cues were presented without their associated intertemporal outcomes (in contrast to the Pavlovian conditioning phase). Therefore, the Pavlovian cue-outcome associations may have extinguished over the course of the transfer phase, potentially weakening the hypothesized influence of Pavlovian cues on instrumental responding over the course of the transfer phase. All Pavlovian cues in the present study were associated with rewards, and reward-predicting cues have in general been shown to increase go-responding (e.g., Huys et al., 2011). Therefore, our hypotheses were formulated in terms of increases in go-responding (and consequently decreases in no-go-responding) as a function of Pavlovian cues (with the strongest hypothesized increase from the large and immediate cues). This hypothesized go-bias implies increased instrumental accuracy on go-trials, and decreased instrumental accuracy on no-go trials. If extinction of Pavlovian associations were to occur over the course of the instrumental phase, however, this go-bias would diminish. We assessed this by investigating whether accuracy on no-go trials increased over the course of the transfer phase (indicating a diminished go-bias). Figure 3C (in the main text) does not show any signs of an increase in accuracy on no-go trials in Experiment 1. Figure 4C (concerning Experiment 2), in contrast, shows a slight divergence in accuracy between go- and no-go trials towards the end of the transfer phase, with an increased accuracy on no-go trials compared to go trials. To investigate this, we ran a model on the no-go trials of the transfer phase, with response (correct/incorrect) as dependent variable, a fixed intercept, and trial number (continuous predictor), Pavlovian reward (immediate large, iL / delayed large, dL / delayed medium, dM / immediate small, iS / delayed small, dS) and their interaction as fixed effects. Random effects included a random intercept and random slopes of trial number, Pavlovian reward, and their interaction as random slopes varying (i) over participants, (ii) over instrumental stimuli (mushrooms), and (iii) over Pavlovian stimuli (square colour); all possible random correlations were also included.

Experiment 1. In line with Figure 3C, there was no overall significant effect of trial number ($b_{\text{TrialNr}} = 0.10$, 95% CI [-0.14, 0.35]), nor was it significant at any of the Pavlovian reward levels ($b_{\text{TrialNr*iL}} = 0.03$, 95% CI [-0.53, 0.59], $b_{\text{TrialNr*dL}} = 0.07$, 95% CI [-0.43, 0.60], $b_{\text{TrialNr*iM}} = 0.06$, 95% CI [-0.50, 0.62], $b_{\text{TrialNr*dM}} = 0.14$, 95% CI [-0.45, 0.67], $b_{\text{TrialNr*iS}} = 0.02$, 95% CI [-0.62, 0.60], $b_{\text{TrialNr*dS}} = 0.31$, 95% CI [-0.79, 1.28]), indicating that instrumental accuracy on no-go trials did not significantly change over the course of the transfer phase.

Experiment 2. Again, we found no overall significant effect of trial number ($b_{\text{TrialNr}} = 0.16$, 95% CI [-0.07, 0.39]), nor was it significant at any of the Pavlovian reward levels

($b_{TrialNr*iL}$ = 0.30, 95% CI [-0.09, 0.73], $b_{TrialNr*dL}$ = 0.14, 95% CI [-0.26, 0.54], $b_{TrialNr*iM}$ = 0.08, 95% CI [-0.34, 0.49], $b_{TrialNr*dM}$ = 0.23, 95% CI [-0.17, 0.65], $b_{TrialNr*iS}$ = 0.04, 95% CI [-0.35, 0.46], $b_{TrialNr*dS}$ = 0.15, 95% CI [-0.49, 0.84]). Thus, in contrast to the visual impression based on Figure 4C, accuracy on no-go trials did not improve over the course of the transfer phase, providing no evidence for extinction effects.

Conclusion

In summary, the performance checks described above showed (i) satisfactory instrumental accuracy at the start of the transfer phase, despite the fact that this phase did not immediately follow the instrumental phase and no outcome feedback was provided, and (ii) no signs of the extinction of Pavlovian cue-outcome associations over the course of the transfer phase, despite these cues being presented without their associated outcomes.

S11. Post-PIT Delay Discounting Results

To assess whether the indifference pairs derived in the pre-PIT delay discounting task remained stable throughout the experiment, we administered a second delay discounting task after the PIT task. The results of this task are described below.

Experiment 1

For four participants, no indifference values could be derived in the post-PIT delay discounting task due to a technical error (n = 4; these participants also experienced technical errors during the pre-PIT task, as described above). For one participant, we failed to derive the indifference value of the immediate small reward because they exclusively selected immediate rewards.[1] In the remaining sample, immediate medium rewards ranged between 1 and 20 cents (M = 12.80, SD = 4.26, Mdn = 13.00, IQR = 6.00), and immediate small rewards ranged between 1 and 16 cents (M = 9.22, SD = 4.09, Mdn = 9.00, IRQ = 7.00).

To investigate the stability of delay discounting from the pre- to the post-PIT delay discounting task, we compared the indifference values derived from both tasks. In order to facilitate the comparison of estimates between Experiment 1 and 2 (which used different reward amounts), the indifference values were converted to discount rates using Mazur's (1987) hyperbolic discounting model. This model holds that $V = A / (1 + kD)$, with V representing the monetary amount of the SS, A the amount of the LL, k the discount rate, and D the delay until delivery of the LL. As D remained constant throughout the task, it was omitted from the model. A constant of 1 was added to each observation to prevent zero-values, allowing the discount rates to be modelled with a lognormal distribution. We ran a model with discount rate as dependent variable, and time (pre-/post-PIT), pair (1/2; with 1 indicating the immediate large versus delayed medium reward pair, and 2 indicating the immediate medium versus delayed small reward pair), and their interaction as predictors. There was a significant effect of time ($b_{\text{PrevsPost}}$ = -0.12, 95% CI [-0.21, -0.02]), with higher discount rates in the post-PIT (estimated k: M = 1.72, Mdn = 1.71) compared to the pre-PIT discounting task (estimated k: $M = Mdn$ = 1.53). The estimated discount rates, as well as the

[1]The inability to derive the second indifference pair could imply that the indifference pairs derived from the pre-PIT task and used in the PIT task were not valid. However, the participant's difference scores in post-PIT liking ratings and go-responding between indifference pair cues were not at either end of the sample distribution. Therefore, following our preregistered criterion, this participant was retained in the sample. Nevertheless, the extreme discounting in the post-PIT task resulted in an immediate medium reward of 1 cent, a k-value of 19, and a strong outlier in our discounting model. As a robustness check, we therefore reran our discounting model while excluding this participant from the sample; conclusions in terms of significant/non-significant results remained unchanged.

observed indifference values ($M_{PrevsPost}$ = 0.70 cents, $Mdn_{\text{PrevsPost}}$ = 1.00 cents) show, however, that the difference across time was small. There was no significant interaction between time and pair ($b_{\text{PrevsPost*1vs2}}$ = -0.04, 95% CI [-0.11, 0.02]), indicating that the effect of time did not differ significantly between the two pairs. In addition, we found no significant main effect of pair (b_{1vs2} = -0.05, 95% CI [-0.11, 0.01]), indicating that discount rates did not differ significantly between pairs.

Experiment 2

Two participants exclusively selected the SS in the second task block of the post-PIT delay discounting task, due to which no indifference values could be derived in this block for these two participants[2]. In the remaining sample, immediate medium rewards subjectively matched to a delayed large reward (indifference pair 1) ranged between €4 and €20 (*M* = 15.92, *SD* = 3.32, *Md* = 16.0, *IQR* = 4.0), and immediate small rewards subjectively matched to a delayed medium reward (indifference pair 2) ranged between €3 and €19 (*M* = 12.54, *SD* = 4.50, *Md* = 13.0, *IQR* = 6.0). Similar to Experiment 1, we converted the indifference values to discount rates (*k*-values) and added a constant of 1 (to avoid zero-values) to investigate whether any significant differences existed between pre- and post-PIT discounting. The statistical model[3] showed no significant effect of time ($b_{\text{PrevsPost}}$ = 0.03, 95% CI [-0.01, 0.06]). Thus, in contrast to Experiment 1, we observed no systematic drift in delay discounting from the pre- to the post-PIT delay discounting task. We did, however, find a significant effect of pair (b_{1vs2} = -0.05, 95% CI [-0.08, -0.01]), with slightly higher discount rates, i.e., more impatient choices, for the second pair (delayed medium versus immediate small; estimated *k*: *M* = *Mdn* = 1.44) compared to the first pair (delayed large versus immediate small; estimated *k: M* = *Mdn* = 1.37). We found no significant interaction between time and pair ($b_{\text{PrevsPost*1vs2}}$ = -0.02, 95% CI [-0.06, 0.02]), showing that the effect of pair was consistent across time points.

Conclusion

In summary, in Experiment 1, we observed a slight but statistically significant increase in intertemporal impatience from the pre-PIT delay discounting task to the post-PIT

[2]Similar to Experiment 1, we inspected these participants' difference scores in post-PIT liking ratings and go-responding between indifference pair cues. As these were not at either end of the sample distribution, we followed our preregistration by retaining these participants in the sample.

[3]We preregistered to run this model with switch points (i.e., the titrator row where the participant switched from choosing the LL to the SS) as dependent variable. However, in order to facilitate comparison of indifference values and model estimates across Experiment 1 and 2 (which used different delay discounting tasks), we used the estimated hyperbolic discount rate as dependent variable. Results from the model on switch points (see S13 for details) were consistent with the results of the discount rate model reported here.

delay discounting task. The increased impatience could suggest that the indifference pairs did not remain stable throughout the experiment. This, in turn, would violate the assumption that by using indifference pairs, we were able to test for PIT effects while controlling for the subjective value of the two rewards forming an indifference pair. However, we wish to point out that the post-PIT liking ratings showed no significant difference in valuation of indifference pair cues, suggesting that, at least until that point in the experiment, the indifference pairs had not changed. Moreover, the direction of the (non-significant) difference in ratings pointed towards a higher valuation of the delayed large instead of the immediate smaller member of the indifference pair. This contrasts with the increased impatience observed in the delay discounting task, which would suggest an increased valuation of the immediate member of the indifference pair. Thus, we consider it unlikely that the indifference pairs changed throughout the task. A more plausible explanation for participants' increased impatience during the post-PIT delay discounting task is that they aimed to shorten the task duration and end the experiment by choosing the immediate reward. By using descriptive rather than experiential delays in Experiment 2, we prevented participants from shortening the task by choosing the immediate reward. The absence of a significant drift in discount rates from the pre- to post-PIT delay discounting task in Experiment 2 supports the idea that the increase in discount rates in Experiment 1 was most likely not due to truly changed indifference pairs, but to a strategy to end the experiment early.

S12. Role of Pavlovian Contingency Awareness

As reported in the main text, in both experiments, several participants scored at or below chance level on the Pavlovian cue-outcome contingency test. Instead of simply excluding these participants from our analyses, we followed previous research (Hogarth et al., 2007; Jeffs & Duka, 2017; Talmi et al., 2008; Trick et al., 2011) by exploring the role of participants' awareness of the Pavlovian contingencies on the reported Pavlovian conditioning and transfer effects. In line with previous studies, we created two new variables, one indicating whether participants scored above (termed *aware* participants) or at or below (termed *unaware* participants) chance level on the Pavlovian contingency test questions that involved the cue-amount associations, and one indicating whether they scored above, or at or below chance level on the Pavlovian contingency test questions that involved the cue-delay associations. We then reran our main Pavlovian and transfer models twice; once while including the cue-amount contingency awareness predictor, and once while including the cue-delay predictor. These predictors were included as fixed effects and random slopes (varying over Pavlovian and instrumental stimuli) in our models. We fitted the most maximal random effects structure that still resulted in an identifiable model given the number of observations per cell.[1]

Below, we report the estimates of our effects of interests for both aware and unaware participants, as well as any significant interaction effects between the awareness predictor and our effects of interest (which would provide statistical evidence for the moderation of our effects of interest by the awareness predictor). It should be noted that the number of participants classified as unaware was low (with the exception of the cue-delay contingency awareness in Experiment 1). Therefore, any non-significant Pavlovian or transfer effects in unaware participants are likely to result from a lack of statistical power instead of (or in addition to) a lack of contingency awareness. Moreover, any significant effects in this group should be interpreted with caution, given the (often extremely) small sample size. Hence, we cannot informatively compare aware and unaware participants. Instead, these analyses are more informative in exploring whether the Pavlovian and transfer results in the subsample

[1]The number of observations per cell was restricted by the cue-outcome contingency awareness variable, as for unaware participants (of which we had relatively few compared to aware participants), some interactions did not have sufficient observations per cell to be modelled as random slopes. Details can be found in the R code available online.

with only aware participants are consistent with the full sample results, or change (e.g., become more pronounced) when the unaware participants are excluded.

Experiment 1

Pavlovian Conditioning

We first included cue-amount contingency awareness to our main Pavlovian model on post-PIT liking ratings (n_{aware} = 45, $n_{unaware}$ = 5). This predictor was modelled as fixed effect (interacting with amount and delay), and as random slope varying over Pavlovian stimuli (interacting with delay). In the subsample consisting of aware participants only (n = 45), findings were consistent with the full sample results, as all amount levels differed significantly from each other (b_{LvsM} = 15.17, 95% CI [6.99, 23.20]; b_{LvsS} = 28.18, 95% CI [16.83, 39.28]; b_{MvsS} = 13.01, 95% CI [3.64, 23.51]), and no significant effect of delay was found (b_{DvsI} = -8.41, 95% CI [-19.40, 1.38]). Moreover, cues associated with indifference pairs did not significantly differ in ratings (b_{DLvsIM} = 9.01, 95% CI [-5.66, 24.33]; b_{DMvsIS} = 6.13, 95% CI [-9.41, 22.22]). In the subsample with unaware participants (n = 5), none of the amount levels differed significantly from each other (b_{LvsM} = 18.40, 95% CI [-1.70, 39.95]; b_{LvsS} = 1.83, 95% CI [-24.19, 29.00]; b_{MvsS} = -16.57, 95% CI [-37.89, 5.06]). A significant interaction indicated that the difference in ratings between medium versus small cues was significantly different between the aware (A) and unaware (U) group ($b_{MvS*UvsA}$ = -29.58, 95% CI [-51.20, -9.84]). No effect of delay was found in the unaware group (b_{DvsI} = -4.46, 95% CI [-16.10, 6.76]), and no significant difference was found between cues associated with members of an indifference pair (b_{DLvsIM} = 20.97, 95% CI [-10.76, 50.78]); b_{DMvsIS} = -1.93, 95% CI [-28.62, 26.60]). A significant interaction showed, however, that the difference in ratings for the second indifference pair was significantly larger in the unaware than the aware group ($b_{DMvsIS*UvsA}$ = -31.07, 95% CI [-60.57, -1.66]). No other significant interactions were found.

Next, we replaced the cue-amount awareness predictor by the cue-delay awareness predictor (n_{aware} = 32, $n_{unaware}$ = 18). This predictor was modelled as a fixed effect (interacting with amount and delay), and as random slope varying over Pavlovian stimuli (interacting with amount and delay, omitting their three-way interaction). Consistent with the full sample results, the aware subsample showed no significant effect of delay (b_{DvsI} = -8.41, 95% CI [-19.40, 1.38]). This suggests that the lack of a delay effect in the full sample was not due to participants' poor cue-delay contingency awareness. However, it should be noted that excluding unaware participants considerably reduced the sample size of the subsample with

aware participants, thereby raising the possibility that the non-significant effect in the aware subsample might be due to a lack of statistical power. Next, the aware subsample showed a significant difference between large versus medium (b_{LvsM} = 14.24, 95% CI [4.31, 24.50]) and between large versus small cues (b_{LvsS} = 27.64, 95% CI [13.37, 41.90], but not between medium versus small cues (b_{MvsS} = 13.40, 95% CI [-0.48, 28.00]). Again, given that the difference between medium and small cues was the smallest effect in the full sample, the reduced sample size in the aware subsample may explain its non-significance. This is supported by the observation that whereas the credible interval for the subsample became wider compared to the full sample CI (possibly due to the low sample size), the point estimate was not smaller in the subsample (full sample: b_{MvsS} = 10.00, 95% CI [0.83, 19.70]). Finally, consistent with the full sample results, the aware subsample showed no significant difference in ratings between cues associated with indifference pairs (b_{DLvsIM} = 6.59, 95% CI [-9.73, 21.67]; b_{DMvsIS} = 4.92, 95% CI [-14.67, 22.44]). The unaware subsample (n = 18) showed no significant effects of delay (b_{DvsI} = -4.46, 95% CI [-16.10, 6.76]). Similar to the aware subsample, there was a significant difference between large versus medium (b_{LvsM} = 15.54, 95% CI [3.20, 27.60]) and between large versus small (b_{LvsS} = 19.75, 95% CI [4.05, 36.50]), but not between medium versus small (b_{MvsS} = 4.21, 95% CI [-11.48, 19.30]) cues. Moreover, there was no significant difference between cues associated with indifference pairs (b_{DLvsIM} = 13.14, 95% CI [-4.68, 31.65]; b_{DMvsIS} = -1.66, 95% CI [-21.69, 18.87]). No significant interactions between the effects of interest (amount, delay, or indifference pairs) and cue-delay contingency awareness were found.

Transfer

First, we added the cue-amount contingency awareness predictor to our transfer model, modelling this predictor as a fixed effect (interacting with amount, delay, and trial type), and as random slope varying over Pavlovian stimuli (interacting with delay and trial type) and instrumental stimuli (interacting with amount, delay, and trial type). In contrast to the full sample results, the aware-only subsample (n = 45) showed no difference in go-responding between large and medium cues. This contrasts with previous literature studying the effect of Pavlovian contingency awareness, which has found transfer effects to be stronger in subsamples including only aware participants (Hogarth et al., 2007; Jeffs & Duka, 2017; Talmi et al., 2008; Trick et al., 2011). It should be noted, however, that the effect in the aware subsample (b_{LvsM} = 0.38, 95% CI [-0.02, 0.76]) was similar in direction and magnitude to the full sample effect (b_{LvsM} = 0.45, 95% CI [0.11, 0.78]), which becomes most clear when evaluating the estimates on the probability scale (full sample: b = 0.61, aware subsample: b =

0.59). Its non-significance in the subsample does suggest, however, that the transfer effect of amount might not be extremely robust. Consistent with the full sample results, the differences between large versus small, and medium versus small were not significant in the aware subsample (b_{LvsS} = 0.47, 95% CI [-0.13, 1.06]; b_{MvsS} = 0.09, 95% CI [-0.43, 0.59]). Moreover, a transfer effect of delay was found (b_{DvsI} = 0.04, 95% CI [-1.27, 1.53]). Finally, cues associated with indifference pairs were not significantly different (b_{DLvsIM} = 0.40, 95% CI [-1.09, 1.85]; b_{DMvsIS} = 0.06, 95% CI [-1.42, 1.55]). Surprisingly, the unaware group (n = 5) did show a significant difference between large and medium cues (b_{LvsM} = 1.06, 95% CI [0.10, 2.03]). However, as described above, this should be interpreted with caution due to the small sample size of this group. The difference between aware and unaware participants for this effect was not significant, as indicated by a non-significant interaction ($b_{LvsM*UvsA}$ = 0.69, 95% CI [-0.32, 1.68]). The unaware group did not show any difference between large versus small, and medium versus small cues (b_{LvsS} = 0.48, 95% CI [-0.99, 2.00]; b_{MvsS} = -0.58, 95% CI [-1.80, 0.56]). Furthermore, no transfer effect of delay was found (b_{DvsI} = 0.39, 95% CI [-1.20, 2.02]), and there was no difference between cues associated with indifference pairs (b_{DLvsIM} = 1.88, 95% CI [-0.01, 3.86]; b_{DMvsIS} = -0.21, 95% CI [-2.21, 1.77]). No significant interactions between the effects of interest and cue-amount contingency awareness were found.

Second, we reran our original transfer model with the cue-delay contingency awareness predictor. The predictor was modelled as a fixed effect (interacting with amount, delay, and trial type), and as random slope varying over Pavlovian stimuli (interacting with delay and trial type, and with amount and trial type) and instrumental stimuli (interacting with amount, delay, and trial type). Consistent with the full sample results, the aware subsample (n = 18) showed no significant effect of delay (b_{DvsI} = 0.06, 95% CI [-0.46, 0.63]). In contrast to the full sample, however, we found no significant effect of amount either (b_{LvsM} = 0.29, 95% CI [-0.18, 0.77]; b_{LsvS} = 0.35, 95% CI [-0.38, 1.04]; b_{MvsS} = 0.06, 95% CI [-0.58, 0.68]). This raises the question why excluding participants that are unaware of the cue-*delay* contingency removes the effect of *amount*. One admittedly speculative explanation, which is further discussed in the general discussion of the main text, is that participants who are unaware of the cue-delay contingencies focus more strongly on the amount attribute associated with the cue (i.e., a fan effect; Anderson, 1974; Anderson & Reder, 1999). Excluding these participants may have therefore weakened the transfer effect of amount. An alternative explanation for the discrepancy in results between the full sample and subsample is that excluding the relatively large group of unaware participants (n = 18) resulted in a loss of

statistical power. It should be noted, however, that the estimate of the effect (b_{LvsM}) also decreased from 0.45 in the full sample to 0.29 in the subsample, which does not support this explanation. In the aware subsample, cues associated with indifference pairs were not significantly different (b_{DLvsIM} = 0.33, 95% CI [-0.46, 1.13]; b_{DMvsIS} = 0.15, 95% CI [-0.70, 1.04]). The unaware subsample (n = 18) also showed no significant effect of delay (b_{DvsI} = 0.03, 95% CI [-0.56, 0.60]), but did show a significant increase in go-responding for large versus medium cues (b_{LvsM} = 0.77, 95% CI [0.16, 1.38]). Although we remain highly cautious in interpreting this finding due to the small sample size of the unaware group, this observation would be in line with the fan effect, in which participants labelled as unaware for cue-delay contingencies show stronger effects for amount. The unaware subsample showed no significant difference between large versus small (b_{LvsS} = 0.69, 95% CI [-0.16, 1.56]) or medium versus small (b_{MvsS} = -0.08, 95% CI [-0.83, 0.68]) cues. Again, cues associated with indifference pairs were not significantly different (b_{DLvsIM} = 0.85, 95% CI [-0.10, 1.78]; b_{DMvsIS} = -0.20, 95% CI [-1.21, 0.76]). No significant interactions between the effects of interest and cue-amount contingency awareness were found.

Experiment 2

Unless specified otherwise, we ran the same models as those specified above for Experiment 1.

Pavlovian Conditioning

Again, we first included cue-amount awareness to our Pavlovian model (n_{aware} = 67, $n_{unaware}$ = 4). In the aware subsample (n = 67), consistent with the full sample results, there was a significant difference in rating between large and small (b_{LvsS} = 20.78, 95% CI [12.39, 29.10]) and between medium and small (b_{MvsS} = 15.20, 95% CI [7.42, 22.60]), but not between large and medium cues (b_{LvsM} = 5.58, 95% CI [-0.78, 11.90]). Furthermore, there was a significant effect of delay (b_{DvsI} = -8.13, 95% CI [-16.20, -1.02]). Cues associated with indifference pairs were not significantly different (b_{DLvsIM} = -2.86, 95% CI [-13.15, 8.01]; b_{DMvsIS} = 7.44, 95% CI [-3.38, 18.99]). In the unaware subsample (n = 4), none of the amount levels differed significantly from each other (b_{LvsM} = -2.26, 95% CI [-23.08, 19.00]; b_{LvsS} = -10.04, 95% CI [-35.31, 17.90]; b_{MvsS} = -7.78, 95% CI [-30.25, 13.50]). The difference between the aware and unaware subsample was significantly different for large versus small and large versus medium cues, as reflected by the interaction coefficients ($b_{LvsS*UvsA}$ = -30.82, 95% CI [-58.50, -4.13]; $b_{MvsS*UvsA}$ = -22.98, 95% CI [-44.80, -1.83]). There was no significant difference between delayed and immediate cues in the unaware group (b_{DvsI} = 0.61, 95% CI [-19.60, 18.84]), and no interaction between the effect of time and the cue-amount contingency

predictor. Again, cues associated with indifference pairs did not significantly differ in ratings (b_{DLvsIM} = 2.78, 95% CI [-25.76, 31.26]; b_{DMvsIS} = 1.40, 95% CI [-27.20, 29.80]), and this did not interact with the cue-amount contingency predictor.

Next, we reran our original transfer model while including the cue-delay awareness predictor to our Pavlovian model (n_{aware} = 62, $n_{unaware}$ = 9). In contrast to the full sample results, the aware subsample showed no significant effect of delay (b_{DvsI} = -7.69, 95% CI [-15.70, 0.52]). As described above, one might expect the effects to become stronger for aware-only subsamples, making this result somewhat unexpected. However, given that the estimate was nearly identical to that found in the full sample, but the CI width had increased compared to the full sample (full sample: b_{DvsI} = -7.63, 95% CI [-13.00, -1.59]), the disparity in results between the full sample and subsample is likely to have resulted from a loss of power due to exclusion of the non-aware participants. Nevertheless, because the number of excluded participants was not extremely high (n = 9), the disparity in results may also suggest that the effect of delay was not very robust. Next, the aware subsample showed a significant difference between large versus small (b_{LvsS} = 20.44, 95% CI [8.86, 31.50]) and between medium versus small cues (b_{MvsS} = 14.22, 95% CI [0.77, 27.40]), but not between large versus medium cues (b_{LvsM} = 6.22, 95% CI [-2.23, 15.30]). These findings are consistent with the full sample results. Finally, the aware subsample showed no significant difference in ratings between cues associated with indifference pairs (b_{DLvsIM} = -9.07, 95% CI [-20.09, 2.00]; b_{DMvsIS} = 7.23, 95% CI [-18.96, 3.72]). The unaware subsample (n = 9) showed no significant effects of delay (b_{DvsI} = -7.31, 95% CI [-21.10, 6.53]) or amount (b_{LvsM} = -2.74, 95% CI [-17.79, 13.30]; b_{LvsS} = -7.61, 95% CI [-12.96, 28.00]; b_{MvsS} = 10.35, 95% CI [-9.00, 30.10]), and no difference between cues associated with indifference pairs (b_{DLvsIM} = -9.08, 95% CI [-30.27, 12.63]; b_{DMvsIS} = 6.08, 95% CI [-17.97, 29.25]). No significant interactions between the effects of interest and cue-delay contingency awareness were found.

Transfer

We first included the cue-amount contingency awareness predictor to our transfer model. For model identifiability reasons, the model differed slightly from that reported above for Experiment 1. That is, we excluded the random slope of the interaction between the cue-amount contingency awareness predictor and trial type varying over instrumental stimuli (as well as all higher-order interactions including this two-way interaction). Consistent with the full sample results, the aware subsample (n = 67) showed no significant effect of amount (b_{LvsM} = 0.12, 95% CI [-0.18, 0.44]; b_{LvsS} = 0.23, 95% CI [-0.25, 0.70]; b_{MvsS} = 0.10, 95% CI [-0.30, 0.53]) or delay (b_{DvsI} = -0.001, 95% CI [-0.48, 0.46]). Cues associated with members

of indifference pairs did not significantly differ in go-responding (b_{DLvsIM} = 0.13, 95% CI [-0.51, 0.77]; b_{DMvsIS} = 0.12, 95% CI [-0.53, 0.82]). The unaware sample (n = 4) also showed no significant effect of amount (b_{LvsM} = -0.36, 95% CI [-1.14, 0.46]; b_{LvsS} = -0.97, 95% CI [-2.09, 0.19]; b_{MvsS} = 0.61, 95% CI [-1.46, 0.31]) or delay (b_{DvsI} = -0.02, 95% CI [-0.98, 0.93]). Again, cues associated with members of an indifference pairs did not significantly differ in go-responding (b_{DLvsIM} = -0.39, 95% CI [-1.71, 0.87]; b_{DMvsIS} = -0.50, 95% CI [-1.79, 0.77]). No significant interactions between the effects of interest and cue-amount contingency awareness were found.

We then reran our transfer model replacing the cue-amount contingency awareness predictor with the cue-delay contingency awareness predictor. This model was identical to that for Experiment 1 (specified above). Consistent with the full sample, the aware subsample (n = 62) showed no significant effect of delay (b_{DvsI} = -0.02, 95% CI [-0.53, 0.47]) or amount (b_{LvsM} = 0.11, 95% CI [-0.25, 0.51]; b_{LvsS} = 0.18, 95% CI [-0.43, 0.75]; b_{MvsS} = 0.07, 95% CI [-0.45, 0.63]). Cues associated with indifference pairs did not have a significantly different effect on go-responding (b_{DLvsIM} = 0.11, 95% CI [-0.56, 0.77]; b_{DMvsIS} = 0.10, 95% CI [-0.68, 0.84]). The unaware subsample (n = 9) also showed no significant effect of delay (b_{DvsI} = -0.02, 95% CI [-0.77, 0.71]) or amount (b_{LvsM} = -0.03, 95% CI [-0.69, 0.63]; b_{LvsS} = -0.17, 95% CI [-1.11, 0.81]; b_{MvsS} = -0.14, 95% CI [-0.88, 0.66]). Again, cues associated with indifference pairs did not have a significantly different effect on go-responding (b_{DLvsIM} = -0.08, 95% CI [-1.10, 0.97]; b_{DMvsIS} = -0.20, 95% CI [-1.27, 0.89]). No significant interactions between the effects of interest and cue-amount contingency awareness were found.

Conclusions

In summary, the subsample analyses for both Experiment 1 and 2 showed largely similar results to the full sample analyses reported in the main text. This suggests that overall, the full sample results were not dependent on participants' Pavlovian cue-outcome contingency awareness. This contrasts with previous research that has found transfer effects to be stronger when participants labelled as unaware were excluded (Hogarth et al., 2007; Jeffs & Duka, 2017; Talmi et al., 2008; Trick et al., 2011). It should again be noted, however, that for most analyses we had a relatively small subsample of unaware participants, preventing us to draw strong conclusions about moderation effects, and about the subsample of unaware participants. However, as described above, the analyses on the aware subsamples allowed us to examine the robustness of the results reported in the main text. This showed that although results were largely consistent with the full sample results, the transfer effect of amount in Experiment 1 was non-significant when unaware participants were excluded, and

the Pavlovian conditioning effect of delay in Experiment 2 was non-significant when participants labelled as delay-unaware were excluded. Several potential explanations were discussed, including a loss of power due to exclusion of the delay-unaware participants, and a fan effect. It should nevertheless be acknowledged that the discrepancy in results between the full sample and subsample suggest that these effects may not be extremely robust.

S13. Switch Point Analyses Delay Discounting Task

For Experiment 2, we preregistered comparing delay discounting between the pre- and post-PIT delay discounting task by using switch points as the dependent variable. In the titrator task used in Experiment 2, participants were given a series of choices presented in rows. Each row presented a choice between a variable immediate reward (sooner-smaller reward; SS) versus a delayed reward (later-larger reward; LL) that was constant across the rows. The row (i.e., choice pair) at which the participants switched from choosing the delayed reward to choosing the immediate reward was termed the *switch point*. For instance, if a participant chose the LL for SS amounts of 2, 4, 6, 8, 10, and 12 euros, but switched to choosing the SS for SS amounts of €14 and larger, they would be assigned a switch point of 7, since the switch from LL to SS was made at the seventh row or choice pair. The task consisted of two blocks, each with one titrator. The reward amount of the immediate member of the indifference pair derived in the first block (e.g., €13) was taken as delayed reward amount in the second block. After the PIT task, participants completed a post-PIT delay discounting task to assess delay discounting stability. The post-PIT delay discounting task was identical to the pre-PIT delay discounting task, except that all rewards were increased by €1 in order to reduce memory effects and prevent participants from simply repeating their choices from the pre-PIT discounting task.

After running Experiment 2, we deviated from our preregistered analysis on switch points in order to facilitate comparisons between Experiment 1 (which used a different delay discounting task and hence did not result in switch points) and Experiment 2. As reported in S11, we ran our main analysis in both experiments on discount rates (*k*-values). Here, we report the results of the preregistered switch point model of Experiment 2. This model included switch point (1-10) as dependent variable, a fixed intercept and time (pre-/post-PIT task), pair (1/2; with 1 indicating the immediate large versus delayed medium reward pair), and their interaction as fixed effects. The random effects included a random intercept, and random slopes of time and pair (omitting their interaction) varying over participants, and all random correlations. A cumulative distribution (with logit link function) was used to account for the ordinal nature of the dependent variable, and posterior predictive checks confirmed the fit of the model with the observed data. All results were consistent with the discount rate model reported in S11. That is, we found no significant effect of time ($b_{\text{PrevsPost}}$ = -1.07, 95% CI [-2.57, 0.31]), indicating that participants did not become more patient or impatient over time. In addition, there was no time by pair interaction ($b_{\text{PrevsPost*1vs2}}$ = 0.20, 95% CI [-1.06,

1.50]). Finally, we did find a significant effect of pair, with lower switch points (and hence more discounting) in the second block (b_{1vs2} = 8.18, 95% CI [6.01, 10.70]). This effect, being stronger than that found in the discount rate model, can at least in part be explained by our task design. That is, the second block often consisted of fewer choices (i.e., rows) than the first block, which consisted of a fixed number of 10 trials, making it likely that this resulted in lower switch points in the second block. Due to this design feature, the preregistered analysis on switch points reported here may not have been appropriate to analyse these data. Therefore, we believe that the analyses on discount rate reported in S11 are better suited to test the stability of discounting across blocks.

References

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, *6*(4), 451–474. https://doi.org/10.1016/0010-0285(74)90021-8

Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, *128*(2), 186–197. https://doi.org/10.1037/0096-3445.128.2.186

Cartoni, E., Balleine, B., & Baldassarre, G. (2016). Appetitive Pavlovian-instrumental Transfer: A review. *Neuroscience and Biobehavioral Reviews*, *71*, 829–848. https://doi.org/10.1016/j.neubiorev.2016.09.020

Cartoni, E., Moretta, T., Puglisi-Allegra, S., Cabib, S., & Baldassarre, G. (2015). The relationship between specific pavlovian instrumental transfer and instrumental reward probability. *Frontiers in Psychology*, *6*, Article 1697. https://doi.org/10.3389/fpsyg.2015.01697

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology*, *67*(2), 319–333. https://doi.org/10.1037/0022-3514.67.2.319

Falk, A., Becker, A., Dohmen, T., Huffman, D., & Sunde, U. (2016). *An experimentally-validated survey module of economic preferences*. IZA Discussion Paper 9674. http://ftp.iza.org/dp9674.pdf

Figner, B., Knoch, D., Johnson, E. J., Krosch, A. R., Lisanby, S. H., Fehr, E., & Weber, E. U. (2010). Lateral prefrontal cortex and self-control in intertemporal choice. *Nature Neuroscience*, *13*(5), 538–539. https://doi.org/10.1038/nn.2516

Hogarth, L., Dickinson, A., Wright, A., Kouvaraki, M., & Duka, T. (2007). The role of drug expectancy in the control of human drug seeking. *Journal of Experimental Psychology:*

Animal Behavior Processes, *33*(4), 484–496. https://doi.org/10.1037/0097-7403.33.4.484

Holmes, N. M., Marchand, A. R., & Coutureau, E. (2010). Pavlovian to instrumental transfer: A neurobehavioural perspective. *Neuroscience and Biobehavioral Reviews*, *34*(8), 1277–1295. https://doi.org/10.1016/j.neubiorev.2010.03.007

Huys, Q. J. M., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R. J., & Dayan, P. (2011). Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS Computational Biology*, *7*(4), Article e1002028. https://doi.org/10.1371/journal.pcbi.1002028

Jeffs, S., & Duka, T. (2017). Predictive but not emotional value of Pavlovian stimuli leads to pavlovian-to-instrumental transfer. *Behavioural Brain Research*, *321*, 214–222. https://doi.org/10.1016/j.bbr.2016.12.022

Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science*, *1*(2), 270–280. https://doi.org/10.1177/2515245918771304

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Vol 5. The effect of delay and of intervening events on reinforcement.* (pp. 55–73). Lawrence Erlbaum Associates.

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt Impulsiveness Scale. *Journal of Clinical Psychology*, *51*(6), 768–774. https://doi.org/10.1002/1097-4679(199511)51:6<768::AID-JCLP2270510607>3.0.CO;2-1

Talmi, D., Seymour, B., Dayan, P., & Dolan, R. J. (2008). Human pavlovian-instrumental transfer. *Journal of Neuroscience*, *28*(2), 360–368. https://doi.org/10.1523/JNEUROSCI.4028-07.2008

Trick, L., Hogarth, L., & Duka, T. (2011). Prediction and uncertainty in human Pavlovian to instrumental transfer. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(3), 757–765. https://doi.org/10.1037/a0022310

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* -- *Fourth Edition*. [Database record]. https://doi.org/10.1037/t15169-000